

Л. Н. Беляева

СЕТЕВЫЕ РЕСУРСЫ В ТЕХНОЛОГИИ ПЕРЕВОДА

Российский государственный педагогический университет им. А. И. Герцена, Российская Федерация, Санкт-Петербург, наб. р. Мойки, 48

В технологической цепочке перевода особое место принадлежит сетевым лексикографическим ресурсам, обеспечивающим ведение терминологических баз языков для специальных целей, выделение не зарегистрированных ранее или новых единиц, их лексикографирование и перевод. Рассматриваются основные особенности современных многоязычных терминологических баз данных, специально анализируются параллельные и сопоставимые корпуса текстов для извлечения терминов и перевода. База терминов EuroTermBank может рассматриваться как опробованная модель многоязычного сетевого ресурса, создание которого должно обеспечить корректную терминологическую и лексикографическую поддержку для перевода документов в различных областях знаний. Рассматриваются методы извлечения терминологии из одноязычных текстов. Библиогр. 24 назв.

Ключевые слова: лингвистические технологии, многоязычные лексикографические ресурсы, банк терминов, средства лингвистической поддержки перевода.

WEB RESOURCES IN TRANSLATION TECHNOLOGY

L. N. Beliaeva

Herzen State Pedagogical University of Russia, 48, Moika emb., Saint Petersburg, Russian Federation

The paper considers the potential of Web resources to be used by various language workers in their practice and studies. Web resources help to support and ensure efficiency, accuracy and correctness of the language worker's results. Main features of the modern multilingual terminological databases, parallel and comparable corpora for terms extraction and translation are under special consideration. Terminological Database EuroTermBank is considered a model for multilingual Web resource, to provide for terminological and lexicographic support for translation of documents in various knowledge domain. Methods of terms extraction from monolingual texts in global English and Russian languages are discussed. Refs 24.

Keywords: linguistic technologies, multilingual lexicographic resources, term bank, tools for linguistic support of translation.

Современный уровень развития лингвистических технологий определяет необходимость уточнения места и функций технического перевода и самого технического переводчика в особой технологической цепочке, включающей системы машинного перевода, комплекс автоматизированных словарей, предметно ориентированный корпус текстов, комплекс прикладных программ. Соответственно, необходимо определить, каковы функции переводчика и терминоведа в новой структуре.

В эпоху развития информационных технологий именно деятельность переводчика является основой извлечения и дальнейшего анализа информации. В такой ситуации особую важность приобретают быстрота выполнения и высокое качество перевода, поскольку перевод, выполненный поздно или некорректно, может привести к критическим последствиям (в сейсмозащите, атомной энергетике, медицине и других областях высокого риска). При современном антропоцентрическом подходе язык научных текстов, объект деятельности переводчика, может рассматриваться как особый функциональный язык, разновидность общелитературного

языка, которому придается большая автономность, — язык для профессиональных (специальных) целей. В отличие от большинства искусственных систем переработки, хранения и передачи информации язык представляет собой открытую динамическую неравновесную метасистему [Пиотровский].

Текст можно рассматривать как результат решения задачи передачи информации и источник (отправную точку) ее извлечения. Соответственно, одним из основных компонентов информационной составляющей специального текста [Герд], т. е. той информации, которую реципиент при определенных условиях и уровне специальных знаний извлекает из текста, является денотативный компонент, связанный с особенностями номинации описываемых в тексте объектов. При включении процесса перевода в систему обработки информации первый этап выявления номинируемых объектов и анализа способов их номинации является задачей терминолога, который выделяет термины в структуре текста и определяет их перевод в рамках конкретного языка для специальных целей (ЯСЦ). В ситуации перевода возможность извлечения информации из научного текста определяется для читателя корректностью передачи терминов — имен объектов. При этом адекватность восприятия текста на лексическом уровне определяется насыщенностью текста именными единицами, степенью компрессии и/или развернутостью номинации объектов.

Резкое изменение ситуации в науке и технике, возникновение новых направлений исследования и, что еще важнее, новых областей знаний приводит сегодня к драматическому отставанию специализированных лингвистических ресурсов, которые необходимы для поддержания исследований и практической работы любого специалиста в области лингвистики и перевода (лингвиста, терминолога, переводчика, лексикографа, составителя технической документации, специалиста в области преподавания языка и т. д.). Эти ресурсы, главным образом сетевые, не только хранятся, но и размечаются в соответствии со стандартом ISO TC 37, разработанным Международной организацией по стандартизации (см., например, проект CLARIN) [Broeder]. В то же время доступные специализированные ресурсы типа универсальных корпусов текстов, грамматик, бумажных словарей или словарей различных автоматизированных систем и систем обработки информации, к сожалению, в своем большинстве не соответствуют ни современному уровню науки и техники, ни основным направлениям развития областей знаний. Эта ситуация определяется не только естественным отставанием таких ресурсов, связанным с необходимостью постоянного терминологического анализа современных текстов, но и с традиционным подходом к созданию и ведению словарей прежде всего на основе уже опубликованных источников и только затем на результатах анализа реальных переведенных текстов. Поскольку именно сетевые терминологические ресурсы являются наиболее используемой переводчиками информацией [Vasiljevs], последующий анализ будет основан на лингвистических технологиях, необходимых для их создания, ведения и поддержания.

Современные инструментальные средства для специалиста, работающего с естественным языком (*language worker*), представляют собой не просто постоянно развивающиеся информационные технологии (ИТ), но основанные на этих технологиях лингвистические ресурсы, ориентированные как на хранение необходимой информации (лексикографической, экспертной, корпусной), так и на системы обработки информации и машинного перевода.

Современное лексикографическое исследование в области создания переводных словарей (как бумажных, так и электронных) предполагает проведение предварительной терминологической работы для отбора и описания терминологии на разных языках, осуществление гармонизации этих описаний и согласование терминологических систем разных языков. Однако в области практической лексикографической работы, т.е. в области создания и ведения пользовательских переводных словарей, что является естественной составляющей профессиональной компетентности специалиста, работающего в сфере перевода, ситуация остается сложной: в различных областях знаний и переводчиками, и специалистами создается огромное количество глоссариев, которые никак не соотносятся друг с другом, и в качестве терминов в них используются словосочетания самой разной длины и структуры.

В процессе работы над переводным словарем принято различать несколько этапов, основными из которых являются:

- создание списка пар терминов конкретной предметной области или под-области (извлечение терминов из параллельных или сопоставимых текстов, их верификация и описание),
- упорядочение пар терминов относительно исследуемого терминополья (систематизация и анализ терминосистемы),
- нормализация пар терминов относительно языка перевода (выбор и утверждение нормативных терминов, унификация, оптимизация),
- кодификация терминосистемы (оформление в виде нормативного словаря, стандартизация и рекомендация терминологии),
- гармонизация терминологии или межязыковое упорядочение [Гринев-Гриневич, с. 19].

Унификация терминов и терминосистем является одним из основных направлений прикладного терминоведения, задачей которого является стандартизация, упорядочение и гармонизация терминологий на различных уровнях описания и фиксации [Лейчик].

Гармонизация терминологии в представленной выше иерархии выполнения лексикографической работы представляет собой заключительную фазу исследования, однако при создании переводного словаря и при опоре на информационные технологии в области использования корпусов текстов этот этап должен реализовываться одновременно с нормализацией. В то же время различие терминосистем разных языков, в частности исходного языка и языка перевода, определяет необходимость установления и изучения именно пар терминов вида *исходный термин — переводной эквивалент*, что позволяет выявить расхождения в терминополье и терминосистемах соответствующей предметной области. Установление таких расхождений и определяет дальнейшее упорядочивание, стандартизацию и унификацию терминологий разных языков, гармонизацию терминосистем этих языков, что, в свою очередь, обеспечивает решение проблем перевода терминов и эффективность межязыковой коммуникации [Беляева и др., 2014].

При этом под термином понимается лексическая единица языка для специальных целей, обозначающая общее — конкретное или абстрактное — понятие теории специальной области знаний или деятельности [Шелов]. Вопрос о соотношении терминов, терминоидов, профессионализмов, жаргонизмов, имен собственных,

номенов и т. п., безусловно важный для терминоведения [Кудашев, с. 100–103], при создании специализированного переводного словаря может не рассматриваться. Дело в том, что, независимо от класса лексической единицы, она должна включаться в переводной словарь, если обладает устойчивостью и достаточно высокой частотой в конкретном языке для специальных целей.

В реальном тексте могут встретиться номинации, характерные для конкретной организации, их значения, как правило, не фиксируются словарями [Кит, с. 150] и представляют особую проблему для адекватного перевода. Так, например, техническая терминология компании «Форд» (внутриорганизационная лексика [Кудашев, с. 106]) представляет собой сложную задачу при выборе эквивалентов перевода для хорошо определенных терминов, используемых в специальном контролируемом языке компании [Rychtycky]. В этом языке существует большое количество терминов, которые описывают процессы автомобилестроения и комплектующие, которые используются только в рамках компании «Форд» по производству автомобилей. Обнаруживается, что многие из терминов не были понятны всем сотрудникам компании, поскольку могут использоваться в рамках только одного подразделения завода (например, *shotgun* в значении ‘механизированное приспособление для развинчивания труб’, а не ‘дробовик’), поэтому должны включаться в соответствующий отраслевой или корпоративный переводной словарь.

Поддержание лексикографической системы, используемой переводчиком, в актуальном состоянии требует постоянного «отслеживания» новых терминов, для чего могут, в частности, использоваться методы собственно лингвистического и лингвостатистического анализа и метрики для выделения терминологических словосочетаний из текста. Автоматическое извлечение терминов (как универбов, так и многокомпонентных лексических единиц — коллокаций) основано на предварительном выравнивании текстов на разных языках, идентификации терминологических единиц в текстах на одном языке и дальнейшем установлении их переводных эквивалентов или, скорее, кандидатов в возможные переводные эквиваленты.

Для выделения терминологических словосочетаний из научных текстов разработано более 80 различных метрик, которые оценивают:

- информацию о сочетаемостных предпочтениях лексических единиц (*unithood*), то есть об их синтагматической близости,
- информацию о степени терминологичности словосочетаний (*termhood*),
- информацию о характерных особенностях (*salience*) словосочетания для конкретного корпуса тестов или терминологии языка для специальных целей [Hertog].

Применение метрик требует создания корпуса текстов и его подготовки для автоматизации терминологического анализа. Идеальным источником материала являются корпуса параллельных текстов, построенные на основе материалов узкой предметной области (статей, монографий, материалов конференций и их переводов на другой язык). Такой корпус должен быть выровнен по предложениям, что позволяет выявлять и анализировать термины и их переводы, оценивать стандартизацию и единство переводов, распространенность конкретных вариантов. На основе корпуса параллельных и сопоставимых текстов можно анализировать номинацию экстралингвистических объектов и особенности терминообразования в условиях различных родных языков авторов текстов на английском языке. Разли-

чия самих исходных текстов, уровней их специализации, целей и профилей конечных пользователей и уровня автоматизации объясняют отсутствие универсальных методов для решения задачи извлечения терминов из текстов.

Методы, используемые для оценки синтагматичности, т. е. устойчивости фиксации элементов словосочетания, основываются на оценках частоты словосочетаний в текстах конкретной предметной области. Для оценки терминологичности словосочетания необходимо сравнение его частотных характеристик в разных предметных областях, т. е. требуется контрастивный подход.

Сегодня под лингвистическими ресурсами, активно применяемыми в лингвистической работе, понимаются естественные или искусственные языки и средства их поддержки, что используется для представления информации о соответствующем естественном языке (словари, онтологии, тезаурусы и пр.), а также для представления ресурсов в системе обработки информации и для решения задач извлечения эмпирической информации. Кроме того, к лингвистическим ресурсам относятся собственно языковые ресурсы (тексты), собранные в мощные базы данных и формирующие источник знаний о языках. Использование лингвистических ресурсов возможно как в «ручном» режиме, так и при решении задач автоматической обработки текста.

Особую часть в технологической цепочке и средствах лингвистической поддержки перевода составляют лексикографические ресурсы, ориентированные на необходимость выполнения терминологической работы. Результаты работы терминолога должны вводиться в систему лексикографических ресурсов до того, как переводчик получает текст и результат машинного перевода. В современной технологической цепочке перевода терминологическая работа не просто является самостоятельным звеном, но осуществляется до собственно перевода [Беляева, 2011].

Следует отметить, что системы, работающие с терминологией, существуют достаточно давно. Еще в 70-х годах XX века крупные компании и правительственные организации создавали машинные языковые фонды: параллельно с экономическим и техническим ростом постоянно появлялась новая терминология, и такие фонды предназначались для унификации терминов, использующихся в текстах на языках для специальных целей и при переводе. Одним из наиболее крупных фондов был банк данных ТЕАМ, разработанный компанией Siemens для работы с европейскими языками, в частности с русским, он включал около 700 000 лексических единиц из различных соответственно сгруппированных тематических областей (естественные науки, бизнес, техника и т. п.) [Hutchins, 2001]. Материалы этого фонда используются и в настоящее время при создании специализированных словарей.

Терминологические базы (или банки) данных (ТБД) представляют собой автоматизированное хранилище терминов. В подобных хранилищах термины снабжены дополнительной информацией как лингвистического (сочетаемость, частотность, принадлежность к семантическому полю), так и экстралингвистического (нормативность, стандартизованность и т. п.) характера. «В зависимости от цели создания ТБД их можно разделить на две группы: ориентированные на обеспечение работ по переводу научно-технической литературы и документации и предназначенные для обеспечения информацией о стандартизованной и рекомендованной терминологии» [Лейчик, с. 284].

В задачу терминологических банков данных входит представление информации об отдельных словах или словосочетаниях (описания, примеры, переводы), эти банки использовались как основа для создания глоссариев специальных текстов и для издания современных специализированных переводных словарей [Hutchins, 1998].

Многие банки данных сразу создавались как многоязычные, почти у всех предусматривался прямой доступ в диалоге, в большинстве банков вводились развернутые описания единиц, являющихся заглавиями словарных статей, некоторые из первых терминологических банков были очень большими. Следует отметить новый подход к работе терминолога и лексикографа: при создании ТБД особое внимание уделялось условиям работы, «дружественности» интерфейса. Новые термины снабжались примерами, текстами на другом языке, дефинициями, полученными из надежных источников, кодами предметных областей и библиографическими ссылками.

Современные многоязычные лексикографические ресурсы по степени универсальности и доступности можно разделить на государственные (например, поддерживаемые Комиссией ЕС) и инициативные, разрабатываемые корпорациями или исследовательскими группами. Рассмотрим далее наиболее активно используемый государственный банк терминологических данных.

Банк данных Eurodicautom [Johnson, Macphail] является самым мощным государственным терминологическим банком, он охватывает все языки Европейского союза и латынь. Европейский союз выпускает законопроекты на 24 языках и работает с 552 комбинациями языков. В основную словарную базу к 2008 г. было включено 1 240 000 словарных статей (5 миллионов терминов) и 325 000 аббревиатур и акронимов. Коды предметных областей основаны на универсальной классификации Леноха [Lenoch]. Пополнение базы данных осуществлялось за счет работы терминологического бюро в Брюсселе и Люксембурге, переводчиков, частных компаний и экспертов в отдельных областях знаний. Обновление системы происходило еженедельно [Rirdance, Vasiljevs]. В этом банке ввод информации был организован в каждом из переводческих бюро в соответствии с их собственными правилами и подходами в зависимости от различных соглашений об использовании и методов сотрудничества каждого языкового сообщества и каждой страны. Поэтому возникла необходимость объединить все отдельные базы данных в единую согласованную базу, разрешающую постоянный ввод материала приблизительно 5000 переводчиков из учреждений ЕС.

В 2008 г. Европейский парламент решил создать самостоятельный орган, в функции которого входит координация, согласование, поддержка и помощь в исследованиях терминологии и сохранении результатов в формате IATE (InterActive Terminology for Europe). Этот формат представляет собой терминологическую реляционную базу данных. Кроме того, в функцию созданного органа входит сотрудничество с переводческими бюро и другими учреждениями при ведении новой базы данных, содержащей миллионы терминов, извлеченных из других баз. Ведение базы предусматривало удаление устаревших и вышедших из употребления терминов и их дубликатов, кроме того, база пополнялась терминологией новых языков. В качестве такого органа Европейским парламентом организован отдел по координации терминологии TermCoord, который осуществляет доступ к термино-

логии ЕС через общедоступный сайт и бесплатные инструментальные средства, а также через Межведомственный терминологический портал EurTerm [Maslias].

Банк данных EuroTermBank представляет собой один из самых мощных государственных терминологических банков, охватывая все языки Европейского союза и латынь. В этом лингвистическом ресурсе объединено 133 локальных ресурса, разработанных в различных бюро перевода ЕС, 2 650 976 терминов (число постоянно увеличивается), 710 705 словарных статей, 221 512 дефиниций на 33 языках. Пополнение словарной базы данных происходит за счет работы терминологического бюро (в Брюсселе и Люксембурге), предложения, поступающие от переводчиков, систематизируются группой Eurodicautom, кроме того, часть информации поступает от частных компаний и экспертов. Обновление системы происходит еженедельно. Структура информации в базе данных EuroTermBank предполагает различные опции выбора исходного языка и языка перевода, предметной области, формы представления информации. При выборе конкретных опций дается информация о вариантах перевода в разных предметных областях и о зафиксированных словосочетаниях. Обращение к словарю осуществляется бесплатно.

Терминологическая база EuroTermBank может рассматриваться как опробованная модель многоязычного сетевого ресурса, создание которого актуально как для языков национальных республик России, так и для языков Таможенного союза ЕАЭС, поскольку может обеспечить корректную терминологическую и лексикографическую поддержку для перевода документов в различных областях сотрудничества и знаний.

Использование ресурса в его современном виде позволяет осуществлять поиск терминов в различных источниках, идентифицировать термины в собственных документах и автоматически извлекать их, просматривать варианты перевода термина в разных предметных областях, искать термины в нескольких языках перевода одновременно, уточнять переводы. Доступ к ресурсу осуществляется непосредственно из Microsoft Word.

В то же время следует иметь в виду, что лексикографические ресурсы, подобные описанным выше, включают главным образом терминологию, извлекаемую в результате стандартизации, и (несмотря на огромные объемы) не способны охватить всю терминологию, особенно для активно развивающихся областей знаний. Сегодня основными недостатками терминологических ресурсов является высокая стоимость и длительное время, необходимое для их создания, недостаточный охват терминологии, особенно для номинации самых современных понятий, недостаточность совместного использования терминологических ресурсов и отсутствие механизмов для вовлечения терминологов-практиков. Следовательно, такие ресурсы должны пополняться системами автоматизации процесса извлечения терминов [Thomas, Atanassova; Cabré].

Метод создания ресурсов, реализуемый с привлечением различных заинтересованных участников, не обязательно специалистов в области лексикографии, терминоведения или перевода, принято называть краудсорсингом (crowdsourcing). Под краудсорсингом понимается процесс, который вовлекает в исследование и сбор информации группу людей (часто называемую толпой — *crowd*), что способствует достижению цели конкретного проекта, поскольку нагрузка распределяется между членами группы [Cibej].

В то же время существуют лингвистические ресурсы, для создания которых привлекаются эксперты самого высокого уровня, развитие таких ресурсов происходит в течение многих лет и накопленная в них информация необходима переводчику. Примером такого ресурса является разработка Принстонского университета, начатая еще в 1984 г., — WordNet, сетевая база данных, представляющая собой очень важный лингвистический ресурс, — по сути, компьютерный тезаурус, размещенный в сети Интернет [Fellbaum]. В этой базе данных слова английского языка (существительные, глаголы, прилагательные и наречия) сгруппированы в наборы когнитивных синонимов (синсеты), при этом каждый синсет номинирует отдельное понятие и является узлом семантической сети. Связи между синсетами определяются концептуальными, семантическими и лексическими отношениями. Система включает 117 000 синсетов, каждый из которых связан с другими синсетами концептуальными отношениями. Кроме того, синсет содержит краткую дефиницию (гlossу) и не менее одного предложения, иллюстрирующего использование его элементов. Многочисленные словоформы представлены в отдельных синсетах, количество которых соответствует числу разных значений.

Наиболее часто встречающимся типом отношений между синсетами является иерархический (гиперонимия, гипонимия или отношение «является экземпляром»). Вторым типом концептуальных отношений между узлами-синсетами является меронимия, отношение типа часть — целое.

Иерархические отношения между синсетами глаголов устроены иначе: глаголы, синсеты которых расположены ближе к основанию деревьев (тропонимы), выражают более специфические способы, характеризующие событие. Конкретный способ зависит от параметров семантического поля; сюда входит объем, скорость или интенсивность. Глаголы, описывающие события, которые обязательно и однонаправленно предполагают друг друга, связаны между собой.

Основным отношением, фиксируемым для прилагательных, является антонимия, пары «непосредственных» антонимов отражают сильную семантическую связь между ними. Для относительных прилагательных указываются производные существительные.

Более подробное описание в терминологии семантических ролей дает другой сетевой ресурс — FrameNet, разрабатываемый в Международном институте информатики в Беркли, Калифорния. Теоретической основой описания является фреймовая семантика, идея которой принадлежит лидеру проекта с его начала в 1997 г. Ч. Филлмору [Atkins, Fillmore].

Семантический фрейм представлен в этом проекте как концептуальная структура, особым формализованным образом описывающая события, отношения, объекты и участников. В базе данных FrameNet содержится около 1200 семантических фреймов, 13 000 лексических единиц (под лексической единицей понимается пара типа слово — дефиниция; многочисленные слова представлены несколькими парами) и более 190 000 предложений, формирующих базу примеров.

Эта лексическая база может использоваться как людьми, так и системами обработки информации. С точки зрения перевода она представляет собой словарь, содержащий более 10 000 значений слов, большая часть сопровождается аннотируемыми примерами, которые демонстрируют значение и употребление. Для исследований в области обработки текстов на естественных языках есть возможность

использовать в качестве эталонного массива более 170 000 предложений, которые аннотированы вручную. Этот массив обеспечивает уникальный набор данных для маркировки семантической роли, используемой в различных приложениях. В системе FrameNet описано более 1000 семантических фреймов и связи между ними, что позволяет соединять более общие и более конкретные фреймы. Поскольку фреймы по сути являются семантическими, они подобны в различных языках, и аналогичные системы разрабатываются сейчас для испанского, немецкого, китайского и японского языков.

Таким образом, можно утверждать, что сетевые лингвистические ресурсы представляют собой важный исследовательский и практический ресурс, использование которого должно быть неотъемлемым условием работы переводчика.

Литература

- Беляева Л. Н. Автоматизированная лексикография: гуманитарные технологии. СПб.: РГПУ им. А. И. Герцена, 2011. 96 с.
- Беляева Л. Н., Данилова О. А., Джепа Т. Л., Камшилова О. Н., Карнун Е. В., Нымм В. Р., Чумилкин С. В. Лексикографический потенциал современных лингвистических технологий. СПб.: Книжный дом, 2014. 168 с.
- Герд А. С. Введение в изучение языков для специальных целей. СПб.: СПбГУ, 2007. 60 с.
- Гринев-Гриневич С. В. Терминоведение. М.: Академия, 2008. 304 с.
- Кит М. С. О стратегии построения высокоэффективных сетевых словарей (на базе разработки словаря LexSite) // Вестник РГГУ. 2010. № 9. С. 149–160.
- Кудашев И. С. Проектирование переводческих словарей специальной лексики. Helsinki: Univ. of Helsinki, 2007. 445 с.
- Лейчик В. М. Прикладное терминоведение и его направления // Прикладное языкознание / под ред. А. С. Герда. СПб.: СПбГУ, 1996. С. 276–286.
- Пиотровский Р. Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы. СПб.: Филол. фак. СПбГУ, 2006. 158 с.
- Шелов С. Д. Еще раз об определении понятия «термин» // Вестник Нижегород. ун-та им. Н. И. Лобачевского. 2010. № 4 (2). С. 795–799.
- Atkins S., Fillmore C. Starting where the dictionaries stop: The challenge for computational lexicography // Computational Approaches to the Lexicon / eds. B. T. S. Atkins, A. Zampolli. Oxford: Oxford Univ. Press, 1994. P. 349–393.
- Broeder D., Kemps-Snijders M., van Uytvanck D., Windhouwer M., Withers P. et al. A Data Category Registry- and Component-based Metadata Framework // Proceedings of “The Seventh International Conference on Language Resources and Evaluation (LREC 2010)”. (Malta, 17–23 May 2010). URL: <http://www.lrec-conf.org/proceedings/lrec2010/index.html> (дата обращения: 16.07.2016).
- Cabré M. T., Estopà R., Vivaldi J. Automatic term detection: A review of current systems // Recent Advances in Computational Terminology / eds. D. Bourigault, C. Jacquemin, M.-C. L’Homme. Amsterdam; Philadelphia: John Benjamins, 2001. P. 53–87.
- Čibej J., Fišer D., Kosem I. The role of crowdsourcing in lexicography // Electronic lexicography in the 21st century: linking lexical data in the digital age: Proceedings of the conference “eLex 2015”. (United Kingdom, Sussex, Herstmonceux Castle, 11–13 August 2015) / eds. I. Kosem, M. Jakubiček, J. Kallas et al. Ljubljana; Brighton Trojina, Institute for Applied Slovene Studies Publ. P. 70–83.
- de Hertog D., Heylen K., Speelman D., Kockaert H. A Variational Linguistic Approach to Term Extraction // Presenting terminology and knowledge engineering resources online: models and challenges: Proceedings of the conference “TKE 2010” / eds. U. Breathnach, F. de Barra Cusack. Dublin: Dublin City Univ. Publ., 2010. P. 226–249.
- Fellbaum C. WordNet and wordnets // Encyclopedia of Language and Linguistics / ed. by K. Brown. 2nd ed. Oxford: Elsevier, 2005. P. 665–670.
- Hutchins J. The Origins of the Translator’s Workstation // Machine Translation. 1998. Vol. 13, iss. 4. P. 287–307.

- Hutchins J. Machine Translation and Human Translation: in Competition or in Complementation? // Machine Translation: Theory & Practice / ed. by M. S. Blekhman. New Delhi: Bahri Publ., 2001. P. 5–20.
- Johnson I., Macphail A. IATE — Inter-Agency Terminology Exchange: Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union // Machine Translation Archive. URL: <http://www.mt-archive.info/LREC-2000-Johnson.pdf> (дата обращения: 07.07.2015).
- Lenoch H. The classification in EURODICATOM and other areas of application // Terminology Bulletin. 1981. Vol. 38. P. 159–178.
- Maslias R. Combining EU Terminology with Communication and Ontology Research // Proceedings of the conference “Terminology and Knowledge Engineering 2014”. (Berlin, 19–21 June 2014). 2014. P. 49–56. URL: http://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf (дата обращения: 18.07.2016).
- Rirdance S., Vasiljevs A. Towards Consolidation of European Terminology Resources: Experience Recommendations from EuroTermBank Project / eds. S. Rirdance, A. Vasiljevs. Riga: Tilde, 2006. 123 p.
- Rychtyckyj N. Machine Translation for Manufacturing: A Case Study at Ford Motor Company // Proceedings of the conference “Eighteenth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-06)” / eds. B. Porter, W. Cheatham. Menlo Park: AAAI Press, 2006. P. 1728–1735.
- Thomas I., Atanassova I. Towards the enrichment of terminological resources by scientific corpora analysis // Electronic lexicography in the 21st century: linking lexical data in the digital age: Proceedings of the conference “eLex 2015”. (United Kingdom, Sussex, Herstmonceux Castle, 11–13 August 2015) / eds. I. Kosem, M. Jakubiček, J. Kallas et al. Ljubljana; Brighton Trojina, Institute for Applied Slovene Studies Publ. P. 136–151.
- Vasiljevs A., Pinnis M., Gornostay T. Service model for semi-automatic generation of multilingual terminology resources // Proceedings of the conference “Terminology and Knowledge Engineering 2014”. (Berlin, 19–21 June 2014). 2014. P. 67–76. URL: http://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf (дата обращения: 18.07.2016).

Для цитирования: Беляева Л. Н. Сетевые ресурсы в технологии перевода // Вестник СПбГУ. Серия 9. Филология. Востоковедение. Журналистика. 2016. Вып. 4. С. 45–55. DOI: 10.21638/11701/spbu09.2016.404.

References

- Atkins S., Fillmore C. Starting where the dictionaries stop: The challenge for computational lexicography. *Computational Approaches to the Lexicon*. Eds. B. T. S. Atkins, A. Zampolli. Oxford, Oxford Univ. Press, 1994, pp. 349–393.
- Beliaeva L. N. *Avtomatizirovannaja leksikografija: gumanitarnye tekhnologii* [Automated lexicography: humanitarian technologies]. St. Petersburg, Herzen State Pedagogical Univ. Press Publ., 2011. 96 p. (in Russian)
- Beliaeva L. N., Danilova O. A., Dzhepa T. L., Kamshilova O. N., Karnup E. V. et al. *Leksikograficheskii potentsial sovremennykh lingvisticheskikh tekhnologii* [Lexicographic potential of modern language technologies]. St. Petersburg, BookHouse LLC Publ., 2014. 168 p. (in Russian)
- Broeder D., Kemps-Snijders M., van Uytvanck D., Windhouwer M., Withers P. et al. A Data Category Registry- and Component-based Metadata Framework. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. Malta, 17–23 May 2010. Available at: <http://www.lrec-conf.org/proceedings/lrec2010/index.html> (accessed: 16.07.2016).
- Cabré M. T., Estopà R., Vivaldi J. Automatic term detection: A review of current systems. *Recent Advances in Computational Terminology*. Eds. D. Bourigault, C. Jacquemin, M.-C. L'Homme. Amsterdam, Philadelphia, John Benjamins Publ., 2001, pp. 53–87.
- Čibej J., Fišer D., Kosem I. The role of crowdsourcing in lexicography. *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015*. Eds. I. Kosem, M. Jakubiček, J. Kallas, S. Krek. United Kingdom, Sussex, Herstmonceux Castle, 11–13 August 2015. Ljubljana; Brighton Trojina, Institute for Applied Slovene Studies Publ., pp. 70–83.
- de Hertog D., Heylen K., Speelman D., Kockaert H. A Variational Linguistic Approach to Term Extraction. *Presenting terminology and knowledge engineering resources online: models and challenges. Proceedings of the TKE 2010*. Eds. U. Breathnach, F. de Barra Cusack. Dublin, Dublin City Univ. Publ., 2010, pp. 226–249.

- Fellbaum C. WordNet and wordnets. *Encyclopedia of Language and Linguistics*. Oxford, Elsevier, 2005, pp.665–670.
- Gerd A. S. *Vvedenie v izuchenie iazykov dlia spetsial'nykh tselei* [Prolegomena to languages for special purposes]. St. Petersburg, St. Petersburg State Univ. Press, 2007. 60 p. (in Russian)
- Grinev-Grinevich S. V. *Terminovedenie* [Terminological science]. Moscow, Academia Publ., 2008. 304 p. (in Russian)
- Hutchins J. The Origins of the Translator's Workstation. *Machine Translation*, 1998, vol. 13, issue 4, pp. 287–307.
- Hutchins J. Machine Translation and Human Translation: in Competition or in Complementation? *Machine Translation: Theory & Practice*. Ed. by M. S. Blekhman. New Delhi, Bahri Publ., 2001, pp. 5–20.
- Johnson I., Macphail A. IATE — Inter-Agency Terminology Exchange: Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union. *Machine Translation Archive*. Available at: <http://www.mt-archive.info/LREC-2000-Johnson.pdf> (accessed 07.07.2015).
- Kit M. S. O strategii postroeniia vysokoeffektivnykh setevykh slovarei (na baze razrabotki slovaria LexSite) [Strategy of creating high-efficiency online dictionaries (on the basis of LexSite dictionary)]. *Vestnik of Russian State University for the Humanities*, 2010, vol. 9, pp. 149–160. (in Russian)
- Kudashev I. S. *Proektirovanie perevodcheskikh slovarei spetsial'noi leksiki* [Designing the specialized translation dictionaries]. Helsinki, Univ. of Helsinki Publ., 2007. 445 p. (in Russian)
- Leichik V. M. Prikladnoe terminovedenie i ego napravleniia [Applied terminological science and its directions]. *Prikladnoe iazykoznanie* [Applied linguistics]. Ed. by A. S. Gerd. St. Petersburg, St. Petersburg State Univ. Press, 1996, pp. 276–286. (in Russian)
- Lenoch H. The classification in EURODICATOM and other areas of application. *Terminology Bulletin*, 1981, vol. 38, pp. 159–178.
- Maslias R. Combining EU Terminology with Communication and Ontology Research. *Proceedings of the conference "Terminology and Knowledge Engineering 2014"*. Berlin, 19–21 June 2014. 2014, pp. 49–56. Available at: http://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf (accessed: 18.07.2016).
- Piotrovskii R. G. *Lingvisticheskaia sinergetika: iskhodnye polozheniia, pervye rezul'taty, perspektivy* [Linguistic synergetics: premise, first results, perspectives]. St. Petersburg, Philological faculty (St. Petersburg State Univ.) Publ., 2006. 158 p. (in Russian)
- Rirdance S., Vasiljevs A. *Towards Consolidation of European Terminology Resources: Experience Recommendations from EuroTermBank Project*. Riga, Tilde Publ., 2006. 123 p.
- Rychtyckij N. Machine Translation for Manufacturing: A Case Study at Ford Motor Company. *Proceedings of the Eighteenth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-06)*. Eds. B. Porter, W. Cheetham. Menlo Park, AAAI Press Publ., 2006, pp. 1728–1735.
- Shelov S. D. Eshche raz ob opredelenii poniatiiia "termin" [Once again on the "term" concept definition]. *Vestnik of Lobachevsky University of Nizhni Novgorod*, 2010, vol. 4 (2), pp. 795–799. (in Russian)
- Thomas I., Atanassova I. Towards the enrichment of terminological resources by scientific corpora analysis. *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015*. Eds. I. Kosem, M. Jakubíček, J. Kallas, S. Krek. United Kingdom, Sussex, Herstmonceux Castle, 11–13 August 2015. Ljubljana; Brighton Trojina, Institute for Applied Slovene Studies Publ., pp. 136–151.
- Vasiljevs A., Pinnis M., Gornostay T. Service model for semi-automatic generation of multilingual terminology resources. *Proceedings of the Terminology and Knowledge Engineering 2014*. Berlin, 19–21 June 2014. 2014, pp. 67–76. Available at: http://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf (accessed: 18.07.2016).

For citation: Beliaeva L. N. Web Resources in Translation Technology. *Vestnik SPbSU. Series 9. Philology. Asian Studies. Journalism*, 2016, issue 4, pp. 45–55. DOI: 10.21638/11701/spbu09.2016.404.

Статья поступила в редакцию 20 января 2016 г.

Статья рекомендована в печать 30 мая 2016 г.

Контактная информация:

Беляева Лариса Николаевна — доктор филологических наук, профессор; lauranbel@gmail.com

Beliaeva Larisa N. — Doctor of Philology, Professor; lauranbel@gmail.com